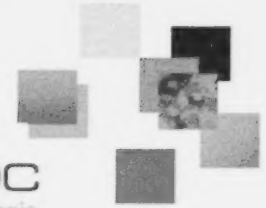
Defence Research and Development Canada | Recherche et développement pour la défense Canada

CAN UNCLASSIFIED



DRDC | RDDC
technologysciencetechnologie

Detecting sociostructural beliefs about group status differences in online discussions

Afzal Upal
DRDC – Toronto Research Centre

Brian Riordan and Heather Wade
Aptima, Inc.

Association for Computational Linguistics

Conference Location: Baltimore, Maryland
Pagination info: 1-6

Date of Publication from Ext Publisher: June 2014

Defence Research and Development Canada
External Literature (N)
DRDC-RDDC-2018-N009
February 2018

CAN UNCLASSIFIED

Canada

IMPORTANT INFORMATIVE STATEMENTS

This document was reviewed for Controlled Goods by Defence Research and Development Canada (DRDC) using the Schedule to the *Defence Production Act*.

Disclaimer: This document is not published by the Editorial Office of Defence Research and Development Canada, an agency of the Department of National Defence of Canada but is to be catalogued in the Canadian Defence Information System (CANDIS), the national repository for Defence S&T documents. Her Majesty the Queen in Right of Canada (Department of National Defence) makes no representations or warranties, expressed or implied, of any kind whatsoever, and assumes no liability for the accuracy, reliability, completeness, currency or usefulness of any information, product, process or material included in this document. Nothing in this document should be interpreted as an endorsement for the specific use of any tool, technique or process examined in it. Any reliance on, or use of, any information, product, process or material included in this document is at the sole risk of the person so using it or relying on it. Canada does not assume any liability in respect of any damages or losses arising out of or in connection with the use of, or reliance on, any information, product, process or material included in this document.

© Her Majesty the Queen in Right of Canada (Department of National Defence), 2018

© Sa Majesté la Reine en droit du Canada (Ministère de la Défense nationale), 2018

Detecting sociostructural beliefs about group status differences in online discussions

Brian Riordan **Heather Wade**
Aptima, Inc.
3100 Presidential Drive
Fairborn, OH 45324
{briordan, hwade}@aptima.com

Afzal Upal
Defence R&D Canada Toronto
1133 Sheppard Ave W
Toronto, ON, M3K 2C9
Afzal.Upal@drdc-rddc.gc.ca

Abstract

Detection of fine-grained opinions and beliefs holds promise for improved social media analysis for social science research, business intelligence, and government decision-makers. While commercial applications focus on mapping landscapes of opinions towards brands and products, our goal is to map "sociostructural" landscapes of perceptions of social groups. In this work, we focus on the detection of views of social group status differences. We report an analysis of methods for detecting views of the legitimacy of income inequality in the U.S. from online discussions, and demonstrate detection rates competitive with results from similar tasks such as debate stance classification.

1 Introduction

Social media and the internet continue to be a vast resource for exploring and analyzing public opinion. While there has been a longstanding focus on detecting sentiment for commercial applications (Liu, 2012), in recent years there has been increased interest in detecting opinions and perspectives in politics and social science more generally (Grimmer & Stewart, 2013). Examples include analyzing people's perceptions of particular political issues by classifying debate stances (Hasan and Ng, 2013) and detecting the expression of ideology (Sim et al., 2013). Research has increasingly turned from detecting opinions and beliefs in general (Prabhakaran et al., 2010) to discerning particular types of opinions or beliefs for specific applications.

The goal of our work is to detect indicators of people's views of social conditions and intergroup perceptions in social media. Working within the framework of Social Identity Theory (Tajfel and Turner, 1979; Tajfel and Turner, 1986; Turner, 1999), we explore detection of the linguistic correlates of sociostructural beliefs. Sociostructural beliefs are abstract theoretical constructs in Social Identity Theory that underpin individual and social identity formation and individual actions that affect the relations between social groups.

For this study, we focus on class-based social groups and the views of individuals on the issue of income inequality. We seek to detect people's views of the legitimacy of the socio-economic structure that has resulted in increasing income inequality, particularly in the U.S. Our approach focuses on comments on news articles related to the issue of income inequality. We develop a series of supervised classifiers for detecting the expression of views on the legitimacy of income inequality. We show promising results comparable to detection rates for other studies of social and political perspectives.

2 Background

Social Identity Theory attempts to account for how subjectively perceived social structure can lead people to define themselves in terms of a shared social identity and thereby produce forms of intergroup behavior. Social identity – how people perceive their relations to the multiple groups to which they belong – is argued to be a crucial part of a person's self-concept. People invoke part of their social identities whenever they think of themselves as belonging to one gender, ethnicity, social class, religion, etc. Group membership and social identity play a role in shaping interpersonal interactions.

Social Identity Theory (as well as social categorization theory) holds that people are sensitive to group status differences and are motivated to view their own social groups positively. These two factors are key drivers of individuals' social identity management strategies. For example, membership

Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media, pages 1–6,
Baltimore, Maryland USA, 27 June 2014. ©2014 Association for Computational Linguistics

in a relatively low-status group may engender perceptions of deprivation, which in turn may result in individuals taking actions to increase their group's status or diminish the status of other groups (Tajfel and Turner, 1979; Tajfel and Turner, 1986). According to Social Identity Theory, a group member's expectations of rewards of group membership are importantly affected by sociostructural beliefs about the nature of group status differences. Group status differences are thought to be shaped by three types of these beliefs:

- *Legitimacy*: the degree to which people believe that group status differences are valid.
- *Stability*: people's sense of how likely the status hierarchy is to last into the future.
- *Permeability*: the perception of how easy it is for outsiders to enter or leave the group.

Based on these sociostructural beliefs and perceptions of the relative deprivation of one's group, people are motivated to take actions to maintain and enhance their group's image.

3 Detecting sociostructural beliefs

A central challenge for extracting sociostructural beliefs is determining where they are likely to occur in natural discourse on the internet. Sociostructural beliefs relate to group status differences – for example, in terms of wealth, power, or prestige. Hence, the most likely context for sociostructural belief expressions is discussions of issues that relate to such social differences.

While debate-focused websites (e.g., createdebate.com, debate.org) hold potential as a data source, we found that in practice such websites had few discussions of issues that might relate to sociostructural beliefs and, furthermore, the number of posts for each topic was generally small. In contrast, we found that highly relevant data can be harvested from comments on news or opinion articles from large newspapers or popular media websites. Articles and op-eds commonly generate hundreds of responses. We considered a variety of topics related to social differences in ethnicity, gender, religion, etc., but found the most data on the topic of income inequality in the U.S. We collected comments across several news articles and op-ed pieces that focused on income inequality.

In the context of income inequality, the social groups are hard to rigorously define, but in comments it was common to observe a dichotomy between "rich" and "poor," or "the 1 percent" and "everyone else". We observed comments on each of the three types of sociostructural beliefs – legitimacy, stability, permeability – but by far the most common topic of discussion was the legitimacy of a large income gap. Therefore, we focused on detecting expressions of legitimacy and leave the extraction of expressions of stability and permeability to future work.

In past survey research related to sociostructural beliefs (Kessler and Mummendey, 2002; Mummendey et al., 1999), participants were asked to respond to explicit statements reflecting sociostructural beliefs – for example, *It is [justified|right|legitimate|accurate|fair] that [proposition]*. However, we found no instances of such explicit expressions in our data. Nevertheless, beliefs about legitimacy are implicit in many instances. For example, consider this comment:

> *Now we are all victims and we should be given our fair share instead of earning our fair share. All the wealth should be redistributed. The wealthy are vilianized. The ones who have been able to rely on their vision, innovation, self motivation, sacrifice and wits are being called out by the envious.Like it or not, the one-percenters are the ones who have advanced humanity to the highest standard ofliving - ever.*

Although there is no explicit articulation of a belief that it is legitimate for income inequality to exist across social groups, for human annotators, it is not difficult to infer that this author likely believes that this is the case. Our goal is to uncover cases like this where sociostructural beliefs are strongly implicit.

We formulated the problem as staged text classification (cf. Lamb et al. (2013)):

1. Finding comments that implicity express the sociostructural belief in the legitimacy or illegitimacy of income inequality (+/-E);
2. Making a binary classification of the author's sociostructural belief (income inequality is legitimate or not) (+/-L).

4 Data Collection

We scraped more than 10,000 comments from articles from major internet media outlets related to

the income inequality issue in the U.S., including *CNN*, *The New York Times*, *Daily Finance*, and marketwatch.com (*The Wall Street Journal*). For example, we collected comments from the CNN op-ed "Is income inequality 'morally wrong'?"[1], which had attracted several thousand comments at the time of data collection (and continues to receive more).

An initial set was randomly selected for annotation for +/-E and +/-L by one of the authors. Another author independently annotated a subset of these comments (N=100) and agreement was assessed. While the agreement was low for the +/-E label (κ = .282), for comments that the annotators agreed were +E, the inter-annotator agreement was high (κ = .916). After the annotators discussed and resolved differences in the +/-E annotation guidelines, the first annotator continued the annotation process to compile a final dataset. Table 1 gives a summary of the final corpus.

	+	-	Total
Expression related to legitimacy (E)	400	1,088	1,488
Support for legitimacy (L)	174	226	400

Table 1: Dataset annotation statistics.

5 Features

5.1 *N*-grams

As with similar tasks such as debate stance classification and sentiment tagging, token-level differences should provide a strong baseline for discriminating between the classes of belief expression (+/-E) and the belief in legitimacy (+/-L). Therefore, we explored a variety of combinations of *n*-gram features, including surface tokens, lemmas, and parts of speech.

5.2 Word classes

Beyond *n*-gram features, we expected that coherent sets of tokens would pattern together for implicit beliefs about legitimacy of status differences. One of the authors coded a total of 24 classes for the income inequality setting based on annotating a subset of about 100 comments. Examples are shown in Table 2. The classes reflected both semantic similarity and, for some, polarity of the sociostructural belief.

[1] http://www.cnn.com/2013/07/25/opinion/sutter-income-inequality-moral-obama/

Word class	**Example words**
income inequality	gap, widening, inequality
lack of income inequality	equal chance, never fair, free society
the non-rich (+)	the 99%, have-nots
the non-rich (+/-)	the poor, middle-class
the non-rich (-)	lazy, dumb
change (+)	fix, make changes
change (-)	redistribution, impose
greed	greed, exploit
hardship	can't afford, cost of living
rich – epithets	shameful, evil, no empathy
poor – epithets	soviet, communist, envy
rich individuals	Buffet, Gates, Bloomberg
society	safety net, playing field
business	companies, profit
money	wealth, income level, salary
the rich (+)	wealthy, those with means
the rich (+/-)	upper middle class
the rich (-)	extreme rich, the 1%
deserve	deserve, earn
work / effort	work harder, effort
success	success, fortune, move up
government	regulation, bloated
taxes	taxes, taxpayer, pay most of
lifestyle	save, budget, responsibility

Table 2: Example word classes.

5.3 Quotation-related features

Excerpts from other posters' comments and quotations of famous individuals are common in our dataset. For example:

> *"Everyone in America has an equal chance an equal opportunity to succeed." Dont know if Id go THAT far.*

The author quotes a previous post's words in order to explicitly disagree with a statement. In this case, *n*-gram features might indicate that the comment should be labeled +L (since comments discussing an "equal opportunity to succeed" typically expressed this belief). However, the second sentence expresses a negation of the ideas in the quoted text. This issue is common in dialogic social media settings, particularly when debating political or social issues, and poses a challenge to surface-oriented classifiers (Malouf and Mullen, 2008). To address this issue, *n*-gram features were computed specifically for text inside

quotes ("quote features") and text outside quotes ("nonquote features"). In the quote above, the words *Everyone in America has an equal chance...* would contribute to the quote n-grams.

6 Experiments

For classification, we experimented with Naive Bayes and MaxEnt (via MALLET[2]) and SVMs (via LIBSVM[3]). Our baseline was a majority class predictor. We began by comparing the results of several different n-gram sets, including n-grams from surface text or lemmatization, binary labels or count features, combinations of unigrams, bigrams, trigrams, and 4-grams, and the inclusion or exclusion of stopwords. We found that the n-grams set of binary labels for unigrams, bigrams, trigrams, and 4-grams after lemmatization had the highest performance. The *inclusion* of stopwords generally afforded better performance; hence we do not remove stopwords.

We explored the hypothesis that this result was due to the inclusion of negation operators among stopwords. Negation may be useful to retain in n-grams to distinguish expressions such as *didn't earn* from *earned*. We removed negation operators from the stopword list. However, other than MaxEnt, performance was worse[4]. What stylometric features that stopwords capture to distinguish authors' beliefs in this task is left for future work.

Classifier	**+/-E**	**+/-L**
MLE	73.1	56.5
MaxEnt	79.9	66.0
Naive Bayes	75.9	**68.3**
SVM	**80.1**	66.3

Table 3: Comparison of classifiers by accuracy on the +/-E and +/-L task with a feature set of: unigram, bigram, trigram, and 4-gram lemma labels, stopwords included. MLE = majority class.

The results for both the +/-E and +/-L tasks are shown in Table 3. We report accuracy following previous related work. We only report results for the staged classifier setting (-E posts were not annotated for +/-L). For the +/-E task, absolute accuracy values were high due to the very unbalanced dataset (cf. Table 1). On the +/-L task, Naive Bayes achieved the highest accuracy score. Our dataset consisted of a mix of short and long comments (M = 45.4 tokens, SD = 37.5 tokens), which, interestingly, was not unfavorable to Naive Bayes (cf. Wang and Manning (2012)). All classifiers were significantly better than the baseline (by paired samples t-tests on accuracy across folds in cross-validation with $p < .05$) in both tasks. On +/-E, MaxEnt and SVM were not significantly different; both performed better than Naive Bayes. On +/-L, there were no significant differences.

Tables 4 and 5 report the results after adding the +/-L problem-specific features to the best n-gram set. The addition of the word class features provides a small improvement in accuracy across the classifiers. MaxEnt's performance approached significance compared to the others ($p < .1$) These results confirm that, for the task of detecting sociostructural beliefs about legitimacy in this domain, words tokens do tend to co-occur in topical and polarity-based word classes. However, it is likely that our word class feature set suffered from limited coverage relative to the diversity of expressions used in the domain.

Feature set	MLE	**ME**	**NB**	**SVM**
n-grams	56.5	66.0	68.3	66.3
+ WC counts	56.5	**70.8**	68.8	67.0
+ WC lab.	56.5	69.5	68.0	67.0
+ WC counts, lab.	56.5	69.8	68.8	66.8

Table 4: Classification accuracies for the +/-L task on variants of word class (WC) feature sets for MaxEnt, NB, and SVM. MLE = majority class.

Table 5 reports the results of adding quotation features. Performance improved with the addition of these features, most notably with the addition of both quote and nonquote features. While these results suggest that accounting for quotations is important, the inclusion of quotation-related features only differentiates between words appearing in quotations from those outside quotations, and does not represent any relationship between the two sets of features. The appearance of terms in a quotation that are typically not found in quotations and that are used by people expressing a particular stance is often a strong indicator that the opinion of the text surrounding the quotation is the opposite of that in the quotation a relationship found by Malouf and Mullen (2008)). Hence, more research that explores relations between terms in and outside of quotations would seem worthwhile.

[2]http://mallet.cs.umass.edu/

[3]http://www.csie.ntu.edu.tw/ cjlin/libsvm/

[4]ME = 66.5, NB = 65.8, SVM = 63.0

Finally, we experimented with combining both word class features and quotation features, but performance did not improve over the results for word class features or quote features alone.

Feature set	MLE	ME	NB	SVM
n-grams	56.5	66.0	68.3	66.3
+ Q count	56.5	67.0	68.3	66.8
+ Q labels	56.5	66.0	68.8	66.3
+ Q count & lab.	56.5	66.5	69.3	65.3
+ NQ labels	56.5	66.3	69.0	65.3
+ Q & NQ	56.5	67.3	70.0	66.3
expl. w/ Q & NQ	56.5	67.3	**70.5**	66.3

Table 5: Classification accuracies for the +/-I task on variants of quotation (Q, NQ) feature sets for MaxEnt, NB, and SVM. MLE = majority class.

7 Error analysis

7.1 Focus on a specific sub-issue

In discussions on income inequality, there are "sub-issues" that are repeatedly discussed in comments, including taxes, welfare, the U.S. economy, and business owners. The difficulty of classifying these kinds of comments stems from the difficulty of deciding whether the comments contain an implicit expression of a sociostructural belief, i.e. the +/- E classification problem. Inference based on world knowledge may be required to chain together the steps that link expressions to beliefs.

7.2 Personal stories used as examples

In discussions involving social status, we observed that people often use personal examples to support their positions.

> *My Dad slept in a dresser drawer on the floor with cotton stuffed under a sheet... He graduated with an engineering degree summa cum laude and has never been un-employed for 45 years because he always worked harder and made himself more valuable than his peers. No GI Bill No Pell Grants No Welfare...*

While a human annotator can usually infer which view on legitimacy such a story supports, the content can seem unrelated to the issue of interest. Similar behavior occurs on debate websites, where descriptions of personal experiences add material irrelevant to stance, often leading to misclassification (Hasan and Ng, 2013).

7.3 Importance of context

While we considered comments independently for our classification task, comments can refer to or reply to previous comments, such that the meaning of a comment can be obscured without the content of these related comments. To address this issue, techniques for incorporating other comments in dialog threads may be fruitful (Walker et al., 2012; Hasan and Ng, 2013).

8 Related Work

The goal of detection of sociostructural beliefs in the context of Social Identity Theory is similar to work in debate stance classification (Anand et al., 2011; Hasan and Ng, 2013; Somasundaran and Wiebe, 2009; Walker et al., 2012). For example, Hasan and Ng (2013) developed methods for classifying author postings on debate websites into binary classes reflecting opposing stances on political issues (e.g., gay marriage). Our setting differs in that "sides" of the issue are only hypothesized (i.e., legitimate/illegitimate) and not given, and stances are never explicitly observed. However, the behavior of posters appears to be similar across debate sites and comments on news articles.

The work here also fits into the increasing focus on content analysis for political and social science analysis (Grimmer and Stewart, 2013). Much recent work has focused on analysis of artifacts from the political arena, such as speeches, floor debates, or press releases (Gerrish and Blei, 2012; Sim et al., 2013; Thomas et al., 2006).

9 Discussion

This work explored the task of detecting latent author beliefs in social media analysis. We focused on the specific problem of detecting and classifying sociostructural beliefs from Social Identity Theory – beliefs about the legitimacy, stability, and permeability of social groups and their status. We collected and analyzed a dataset of social media comments centering on the issue of income inequality and sought to classify implicit author beliefs on the legitimacy of class-based income disparity. Because of the heavily implicit nature of sociostructural belief expression, we formulated the detection problem as a form of text classification. Our approach achieved classification accuracies competitive with results from similar tasks such as debate stance classification.

References

Anand, Pranav, Walker, Marilyn, Abbott, Rob, Tree, Jean. E. Fox, Bowmani, Robeson, and Minor, Michael. 2011. Classifying stance in online debate. In *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*.

Gerrish, Sean M., and Blei, David M. 2012. How They Vote: Issue-Adjusted Models of Legislative Behavior. In *Advances in Neural Information Processing Systems*.

Grimmer, Justin, and Stewart, Brandon M. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267297.

Hasan, Kazi Saidul, and Ng, Vincent. 2013. Frame Semantics for Stance Classification. *CoNLL-2013*, 124.

Kessler, Thomas, and Mummendey, Amélie. 2002. Sequential or parallel processes? A longitudinal field study concerning determinants of identity-management strategies. Journal of Personality and Social Psychology, 82(1), 75-88.

Lamb, Alex, Paul, Michael J., and Dredze, Mark. 2013. Separating fact from fear: Tracking flu infections on Twitter. In *Proceedings of NAACL-HLT*.

Liu, Bing. 2012. *Sentiment analysis and opinion mining*. Morgan & Claypool.

Malouf, Robert, and Mullen, Tony. 2008. Taking sides: User classification for informal online political discourse. *Internet Research*, 18(2), 177-190.

Mummendey, Amélie, Klink, Andreas, Mielke, Rosemarie, Wenzel, Michael, and Blanz, Mathias. 1999. Sociostructural characteristics of intergroup relations and identity management strategies: results from a field study in East Germany. *European Journal of Social Psychology*, 29(2-3), 259285.

Prabhakaran, Vinodkumar, Rambow, Owen, and Diab, Mona. 2010. Automatic committed belief tagging. In *Proceedings of COLING*.

Sim, Yanchuan, Acree, Brice, Gross, Justin H., and Smith, Noah A. 2013. Measuring ideological proportions in political speeches. In *Proceedings of EMNLP*.

Somasundaran, Swapna, and Wiebe, Janyce. 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*.

Tajfel, Henri and Turner, John C. 1979. An integrative theory of intergroup conflict. In W.G. Austin and S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 3347). Monterey, CA: Brooks-Cole.

Tajfel, Henri and Turner, John C. 1986. The social identity theory of intergroup behaviour. In S. Worchel, and W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 724). Chicago, IL: Nelson-Hall.

Thomas, Matt, Pang, Bo, and Lee, Lillian. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of EMNLP*.

Turner, John C. 1999. Some current issues in research on social identity and self-categorization theories In Ellemers, N., Spears, R., Doosje, B. *Social identity* (pp. 6-34). Oxford: Blackwell.

Walker, Marilyn A., Anand, Pranav, Abbott, Robert, and Grant, Ricky. 2012. Stance classification using dialogic properties of persuasion. In *Proceedings of NAACL-HLT*.

Wang, Sida I., and Manning, Christopher D. 2012. Baselines and Bigrams: Simple, Good Sentiment and Topic Classification. In *Proceedings of ACL*.

DOCUMENT CONTROL DATA

*Security markings for the title, authors, abstract and keywords must be entered when the document is sensitive

1. ORIGINATOR (Name and address of the organization preparing the document. A DRDC Centre sponsoring a contractor's report, or tasking agency, is entered in Section 8.)

DRDC – Toronto Research Centre
Defence Research and Development Canada
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9
Canada

2a. SECURITY MARKING (Overall security marking of the document including special supplemental markings if applicable.)

CAN UNCLASSIFIED

2b. CONTROLLED GOODS

NON-CONTROLLED GOODS
DMC A

3. TITLE (The document title and sub-title as indicated on the title page.)

Detecting sociostructural beliefs about group status differences in online discussions

4. AUTHORS (last name, followed by initials – ranks, titles, etc., not to be used)

Upal, A.; Riordan, B.; Wade, H.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total pages, including Annexes, excluding DCD, covering and verso pages.)	6b. NO. OF REFS (Total references cited.)
February 2018	6	18

7. DOCUMENT CATEGORY (e.g., Scientific Report, Contract Report, Scientific Letter.)

External Literature (N)

8. SPONSORING CENTRE (The name and address of the department project office or laboratory sponsoring the research and development.)

DRDC – Toronto Research Centre
Defence Research and Development Canada
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9
Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. DRDC PUBLICATION NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC-RDDC-2018-N009	

11a. FUTURE DISTRIBUTION WITHIN CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

Public release

11b. FUTURE DISTRIBUTION OUTSIDE CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

12. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Use semi-colon as a delimiter.)
Information Extraction; Social Media; Machine Learning
13. ABSTRACT/RESUME (When available in the document, the French version of the abstract must be included here.)